

Joint Transmitter and Receiver Optimization for Improper-Complex Second-Order Stationary Data Sequence

Jeongho Yeo, Joon Ho Cho[†], and James S. Lehnert

Abstract

In this paper, the transmission of an improper-complex second-order stationary data sequence is considered over a strictly band-limited frequency-selective channel. It is assumed that the transmitter employs linear modulation and that the channel output is corrupted by additive proper-complex cyclostationary noise. Under the average transmit power constraint, the problem of minimizing the mean-squared error at the output of a widely linear receiver is formulated in the time domain to find the optimal transmit and receive waveforms. The optimization problem is converted into a frequency-domain problem by using the vectorized Fourier transform technique and put into the form of a double minimization. First, the widely linear receiver is optimized that requires, unlike the linear receiver design with only one waveform, the design of two receive waveforms. Then, the optimal transmit waveform for the linear modulator is derived by introducing the notion of the propriety frequency function of a discrete-time

J. Yeo and J. S. Lehnert are with the School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN 47907-2035 (e-mail: {jeonghoyeo, lehnert}@purdue.edu).

J. H. Cho is with the Department of Electrical Engineering, Pohang University of Science and Technology (POSTECH), Pohang, Gyeongbuk 790-784, Korea (e-mail: jcho@postech.ac.kr).

[†] corresponding author

random process and by performing a line search combined with an iterative algorithm. The optimal solution shows that both the periodic spectral correlation due to the cyclostationarity and the symmetric spectral correlation about the origin due to the impropriety are well exploited.

Index Terms

Cyclostationarity, improper-complex, joint transmitter and receiver optimization, mean-squared error (MSE), vectorized Fourier transform (VFT).

I. INTRODUCTION

An information-bearing signal encountered in communications and signal processing often exhibits periodicity in its mean and auto-covariance functions and thus it is well modeled by a wide-sense cyclostationary (WSCS) random process [1]. This structure in the first-order and the second-order statistics has long been exploited in the design of many communications and signal processing systems [2], [3].

One of the classical problems related to the processing of WSCS random processes is a joint optimization of the transmitter (Tx) and receiver (Rx) in a communication system. In [4]–[7], real-baseband pulse amplitude modulation (PAM) of a wide-sense stationary (WSS) real-valued data symbol sequence is considered with a linear Rx for use over an additive WSS colored noise channel. Under the minimum mean-squared error (MMSE) optimality criterion and the average transmit power constraint, the jointly optimal transmit and receive waveforms are derived. It is shown that, interestingly, the waveforms have nonzero spectral values only on a generalized Nyquist interval [6] with length equal to the minimum bandwidth required to satisfy the Nyquist condition for zero intersymbol interference (ISI) [1].

This joint optimization problem is extended in [8] to complex-baseband quadrature amplitude modulation (QAM) of a WSS complex-valued data symbol sequence. Under the linear MMSE (LMMSE) optimality criterion and the average transmit power constraint, the jointly optimal transmit and receive waveforms are derived for use over an additive WSCS noise channel. It is well known that a WSCS noise model is better than a WSS model for the case in which data-like QAM interferences are present as well

as an ambient Gaussian noise [1]. In contrast to the previous results only with an additive WSS noise, the optimal waveforms are shown in general to have nonzero spectral values on a frequency interval whose length is greater than that of the generalized Nyquist interval. This is because, unlike a WSS random process, a WSCS random process possesses non-zero correlation in the frequency domain among the components that are spaced integer multiples of the symbol rate apart [9]. To exploit such spectral correlation of the WSCS random process, a vectorized Fourier transform (VFT) technique is employed in [8]. This technique is motivated by the harmonic series representation [9] of a WSCS random process, and the use of that representation for joint Tx and Rx optimizations in cyclostationary interference and noise has been examined in [10] and [11].

The results in [8], [10], [11], however, have considered only the real passband or, equivalently, the complex baseband transmission of a proper-complex data sequence. Hence, these results are not directly applicable to, e.g., the real passband transmission of a BPSK data sequence, which is an improper-complex data sequence in complex baseband. Recall that complex-valued random variables, vectors, and processes are called proper if their complementary covariance, complementary covariance matrix, and complementary auto-covariance function (a.k.a. pseudo-covariance, pseudo-covariance matrix, and pseudo-covariance function) vanish, respectively [12]. Otherwise, they are called improper [13]. Although the complex envelopes of the majority of digitally modulated signals are proper, there still remain other digitally modulated signals whose complex envelopes have non-vanishing complementary auto-covariance functions [13]. For example, the complex envelopes of PAM, vestigial sideband PAM, unbalanced QAM, offset quaternary phase-shift keying (OQPSK), and Gaussian minimum shift keying are improper.

Among these improper-complex signals, we focus in this paper on a linear modulation of an improper-complex data sequence using only one transmit waveform. In particular, we consider an improper-complex data sequence that is well modeled by a zero-mean improper-complex second-order stationary (SOS) random process for which the auto-covariance and the complementary auto-covariance functions depend only on the time difference [13]. This results in an improper-complex second-order cyclostationary (SOCS) transmitted signal. For example, PAM, vestigial sideband PAM, and unbalanced QAM fall into this

category. It is assumed that such an improper-complex SOCS signal is transmitted over a strictly band-limited frequency-selective linear time-invariant (LTI) channel whose output is corrupted by an additive proper-complex SOCS random process. As already mentioned, proper-complex SOCS random processes well model the complex envelopes of the majority of digitally modulated signals as well as the complex envelope of an additive Gaussian noise.

Our objective is to extend the aforementioned joint optimizations of the Tx and Rx for proper-complex WSCS signaling to a joint Tx and Rx optimization problem for improper-complex SOCS signaling under the MMSE optimality criterion and the average transmit power constraint. It is well known that the second-order properties of an improper-complex signal are not well captured by a linear Rx, but instead by a class of nonlinear Rx's called widely linear Rx's [13]. There are two types of widely linear Rx's. The first one linearly processes the signal augmented by its complex conjugate, whereas the second one linearly processes the real part of the signal augmented by the imaginary part. In this paper, the first type of widely linear processing also referred to as the linear-conjugate linear (LCL) filtering [14] is employed. It is noteworthy that, unlike the joint optimizations in [8], [10], [11], we now need to find two receive waveforms under the widely linear MMSE (WLMMSE) optimality criterion, where one is employed to filter the complex envelope of the received signal and the other to filter its complex conjugate.

The VFT technique again enables us to convert the objective function and the average transmit power constraint described initially in the time domain into those in the frequency domain. Unlike the previous joint optimizations, the objective function is now expressed in terms of the VFT of the transmit waveform augmented by the VFT of its complex conjugate and the VFT of a receive waveform augmented by the VFT of the other receive waveform. Using these augmented vector-valued functions, we derive the optimal waveforms of the WLMMSE Rx in a straightforward way as a function of the transmit waveform. It is shown that the two receive waveforms of the WLMMSE Rx exploit not only the periodic spectral correlation due to the cyclostationarity, but also the symmetric spectral correlation about the origin due to the impropriety [13].

To derive the optimal transmit waveform, we devise the notion of the impropriety frequency function

of the transmitted improper-complex SOS data sequence by using the relation between the power spectral density (PSD) and the complementary PSD of the random process. This real-valued non-negative function converts the transmit waveform optimization problem into an equivalent convex optimization problem to find the optimal energy density of the transmit waveform. Then, a line search combined with an iterative algorithm is proposed to solve the problem. After finding the optimal energy density, the optimal transmit and receive waveforms are obtained. Numerical results provide an example of joint waveform design and also show the effect of the propriety frequency function on the mean-squared error (MSE) performance.

The rest of this paper is organized as follows. In Section II, the system model is described and the problem is formulated in the time domain. In Section III, the problem is reformulated in the frequency domain. In Section IV, the propriety frequency function is introduced and the jointly optimal transmit and receive waveforms are derived. Numerical results are provided in Section V, and concluding remarks are offered in Section VI.

II. SYSTEM MODEL AND PROBLEM FORMULATION

In this section, we describe the system model and formulate the optimization problem in the time domain. The system model is an extension of that in [8], which only considers the transmission and reception of a proper-complex SOS data sequence, to now allow improper-complex SOS sequences. The optimality criterion of the joint optimization problem is also extended from the LMMSE criterion to the WLMSE criterion.

A. System Model

A Tx and an Rx operate over a real passband to transmit a data sequence $\{b[l]\}_{l \in \mathbb{Z}}$. Fig. 1 shows the system block diagram in complex baseband. The data sequence $\{b[l]\}_{l \in \mathbb{Z}}$ is assumed well modeled by a zero-mean improper-complex SOS random process with auto-covariance and complementary auto-covariance functions given, respectively, by $m[k] \triangleq \mathbb{E}\{b[k+l]b[l]^*\}$ and $\tilde{m}[k] \triangleq \mathbb{E}\{b[k+l]b[l]\}$, where the superscript $*$ denotes complex conjugation. By applying the discrete-time Fourier transform (DTFT) operations to $m[k]$ and $\tilde{m}[k]$, the PSD $M(f)$ and the complementary PSD $\tilde{M}(f)$ of the data sequence $\{b[l]\}_{l \in \mathbb{Z}}$ are derived, respectively, as $M(f) \triangleq \sum_{k=-\infty}^{\infty} m[k]e^{-j2\pi f k}$ and $\tilde{M}(f) \triangleq \sum_{k=-\infty}^{\infty} \tilde{m}[k]e^{-j2\pi f k}$.

The Tx to be designed employs linear modulation with symbol transmission rate $1/T$ [symbols/sec], where the transmit waveform is denoted by $s(t)$. The transmitted signal $\sum_{k=-\infty}^{\infty} b[k]s(t - kT)$ is passed through a strictly band-limited channel that is modeled by an LTI system with impulse response $h(t)$ having the one-sided bandwidth B [Hz] in complex baseband.

The received signal denoted by $Z(t)$ consists of the signal from the Tx and an additive interference-plus-noise signal $N(t)$, where the latter is modeled by a zero-mean proper-complex SOCS random process with fundamental cycle period T_0 . It is assumed that the multiplicative inverse T of the symbol transmission rate of the desired signal is chosen as an integer multiple of T_0 . Thus, $Z(t)$ can be written as

$$Z(t) = \sum_{k=-\infty}^{\infty} b[k]p(t - kT) + N(t), \quad (1)$$

where $p(t) \triangleq h(t) * s(t)$ denotes the overall response with the operator $*$ denoting the convolution integral. There should be no confusion from the superscript $*$ that denotes the complex conjugation.

In (1), it can be easily shown that the desired signal component $X(t) \triangleq \sum_{k=-\infty}^{\infty} b[k]p(t - kT)$ becomes a zero-mean SOCS random process due to the second-order property of the zero-mean SOS data sequence $\{b[l]\}_{l \in \mathbb{Z}}$. In other words, the mean, the auto-covariance, and the complementary auto-covariance functions of $X(t)$ satisfy, respectively, $\mu_X(t) \triangleq \mathbb{E}\{X(t)\} = 0$, $r_X(t, s) \triangleq \mathbb{E}\{X(t)X(s)^*\} = r_X(t + T, s + T)$, and $\tilde{r}_X(t, s) \triangleq \mathbb{E}\{X(t)X(s)\} = \tilde{r}_X(t + T, s + T)$, $\forall t, \forall s$. In what follows, we also call $r_X(t, s)$ and $\tilde{r}_X(t, s)$ the auto-correlation and the complementary auto-correlation functions, respectively, because $X(t)$ has mean zero.

In (1), it can be straightforwardly shown that the interference-plus-noise signal $N(t)$ is SOCS with mean zero and cycle period T , because T is assumed to be an integer multiple of T_0 , i.e., $\mu_N(t) \triangleq \mathbb{E}\{N(t)\} = 0$, $r_N(t, s) \triangleq \mathbb{E}\{N(t)N(s)^*\} = r_N(t + T, s + T)$, and $\tilde{r}_N(t, s) \triangleq \mathbb{E}\{N(t)N(s)\} = \tilde{r}_N(t + T, s + T)$, $\forall t, \forall s$. Now that $Z(t)$ is a summation of two uncorrelated zero-mean SOCS random processes with cycle period T , it is also a zero-mean SOCS random processes with cycle period T .

It is well known [13] that, for a vector-valued signal model, a widely linear Rx employing two linear filters outperforms a linear Rx employing only one linear filter when either the desired signal or the interference-plus-noise signal is improper. Thus, in this paper, we employ two LTI filters with

impulse responses $w_1(-t)^*$ and $w_2(-t)^*$ to process the improper-complex SOCS process $Z(t)$ and its complex conjugate $Z(t)^*$, respectively. The two LTI filters are followed by uniform samplers with rate $1/T$ [samples/sec], and then the sequence of decision statistics $\{z[l]\}_{l \in \mathbb{Z}}$ is obtained as the sum of the sampler outputs, i.e.,

$$z[l] \triangleq z_1[l] + z_2[l], \quad (2)$$

where the sampler outputs $z_1[l]$ and $z_2[l]$ are defined, respectively, as

$$z_1[l] \triangleq w_1(-t)^* * Z(t) \Big|_{t=lT} = \int_{-\infty}^{\infty} w_1(t - lT)^* Z(t) dt \quad \text{and} \quad (3a)$$

$$z_2[l] \triangleq w_2(-t)^* * Z(t)^* \Big|_{t=lT} = \int_{-\infty}^{\infty} w_2(t - lT)^* Z(t)^* dt. \quad (3b)$$

B. Problem Formulation in Time Domain

Our objective is to find the transmit and receive waveforms $s(t)$, $w_1(t)$, and $w_2(t)$ that jointly minimize the MSE given by

$$\varepsilon(s(t), w_1(t), w_2(t)) \triangleq \mathbb{E}\{|b[l] - z[l]|^2\}, \quad (4)$$

where $s(t)$, $w_1(t)$, and $w_2(t)$ are the parameters to be designed. Since T is an integer multiple of the fundamental cycle period T_0 of the interference-plus-noise signal, it can be easily shown that the MSE defined in (4) as the objective function of the optimization problem is the same regardless of the value of l .

The average transmit power constraint is then imposed on this joint optimization problem. Since the transmitted signal is SOCS with cycle period T , the average transmit power \bar{P} can be defined as

$$\bar{P} \triangleq \mathbb{E} \left\{ \frac{1}{T} \int_{\langle T \rangle} \left| \sum_{k=-\infty}^{\infty} b[k] s(t - kT) \right|^2 dt \right\}, \quad (5)$$

where $\langle T \rangle$ denotes any integration interval of length T [sec]. Thus, the constraint is given by $\bar{P} = P_T$ for some $P_T > 0$. Therefore, the joint optimization problem is given by

$$\text{minimize}_{(s(t), w_1(t), w_2(t))} \varepsilon(s(t), w_1(t), w_2(t)) \quad (6a)$$

$$\text{subject to} \quad \bar{P} = P_T. \quad (6b)$$

Problem 1:

III. PROBLEM REFORMULATION IN FREQUENCY DOMAIN

In this section, Problem 1 described in the time domain is reformulated in the frequency domain. To proceed, we first review the notions of the VFT and the matrix-valued PSD. Then, by proposing the notion of the matrix-valued complementary PSD and the methods to augment the VFTs of the transmit and receive waveforms, we convert the objective function (4) and the average transmit power constraint (5) to equivalent expressions in the frequency domain.

A. Review of VFT and Matrix-Valued PSD

In this subsection, we briefly review the notions of excess bandwidth, the Nyquist interval, the VFT, and the matrix-valued PSD. For details, see [8].

Given a pair $(B, 1/T)$ of a bandwidth and a reference rate, the excess bandwidth β is defined as $\beta \triangleq 2BT - 1$ and the Nyquist interval \mathcal{F} is defined as $\mathcal{F} \triangleq \{f : -\frac{1}{2T} \leq f < \frac{1}{2T}\}$.

Given a pair $(B, 1/T)$ and a deterministic function $p(t)$ having the continuous-time Fourier transform (CTFT) $P(\xi) \triangleq \int_{-\infty}^{\infty} p(t)e^{-j2\pi\xi t}dt$, the VFT $\mathbf{p}(f)$ of $p(t)$ is defined as a vector-valued function of $f \in \mathcal{F}$ that is equivalent to $P(\xi)$. In particular, the k th entry of $\mathbf{p}(f)$ is given by $[\mathbf{p}(f)]_k \triangleq P(f + \frac{k-L-1}{T})$ for $k = 1, 2, \dots, 2L+1$, where $L \triangleq \lceil \beta/2 \rceil$.

Given a pair $(B, 1/T)$ and an SOCS random process $N(t)$ with cycle period T having the auto-correlation function $r_N(t, s)$, the matrix-valued PSD $\mathbf{R}_N(f)$ of $N(t)$ is defined as a matrix-valued function of $f \in \mathcal{F}$, whose (k, l) th entry is given by $[\mathbf{R}_N(f)]_{k,l} \triangleq R_N^{(k-l)}(f + (l - L - 1)/T)$ for $k, l = 1, 2, \dots, 2L+1$, where $R_N^{(k)}(\xi)$ is the CTFT of $r_N^{(k)}(\tau)$ that is obtained by applying the Fourier series expansion to $r_N(t, t - \tau)$, i.e., $r_N(t, s) = \sum_{k=-\infty}^{\infty} r_N^{(k)}(t - s)e^{j2\pi kt/T}$.

In using the above definitions, it is assumed that the parameter B is chosen as bandwidth in complex baseband over which the Rx can observe and process a signal and that the parameter $1/T$ is chosen as the symbol transmission rate of the Tx. It is also assumed that the frequency band over which the Tx can emit non-zero power is identical to the frequency band of the Rx. For a general case where these two frequency bands are different, the notion of virtual legacy Rx's and the orthogonal constraint at the virtual legacy Rx's can be employed as is done in [15] for the transmission of a proper-complex data sequence.

Due to the above assumption on the frequency band that can be used by the Tx and the Rx, the first and the last entries of the VFT of the transmit waveform need to be always zero for $-1/(2T) \leq f \leq L/T - B$ and $B - L/T \leq f \leq 1/(2T)$, respectively. For this, the notion of the effective VFT is employed as discussed in [8], [15], and [16]. The effective VFT is defined as a variable-length vector-valued function of $f \in \mathcal{F}$ by removing the first and the last entries of the VFT for $-1/(2T) \leq f \leq L/T - B$ and $B - L/T \leq f \leq 1/(2T)$, respectively. In what follows, the length of the effective VFT is denoted by

$\mathcal{N}(f)$. For details, see [16, Eq. (14)]. Similarly, the effective matrix-valued PSD can be also defined as an $\mathcal{N}(f)$ -by- $\mathcal{N}(f)$ matrix-valued function of $f \in \mathcal{F}$ by removing both the first row and column of the matrix-valued PSD for $-1/(2T) \leq f \leq L/T - B$ and by removing both the last row and column for $B - L/T \leq f \leq 1/(2T)$.

B. Problem Reformulation in Frequency Domain

In this subsection, the objective function and the average transmit power constraint in Problem 1 are converted into equivalent expressions in the frequency domain. To begin with, we propose the notion of the matrix-valued complementary PSD of an improper-complex SOCS random process.

Definition 1: Given a pair $(B, 1/T)$ and an improper-complex SOCS random process $X(t)$ with cycle period T and complementary auto-correlation function $\tilde{r}_X(t, s)$, let $\tilde{R}_X^{(k)}(\xi)$ be the CTFT of $\tilde{r}_X^{(k)}(\tau)$ that is obtained by applying the Fourier series expansion to the periodic signal $\tilde{r}_X(t, t - \tau) = \tilde{r}_X(t + T, t + T - \tau)$, $\forall t$, i.e., $\tilde{r}_X(t, s) = \sum_{k=-\infty}^{\infty} \tilde{r}_X^{(k)}(t - s) e^{j2\pi kt/T}$. Then, the matrix-valued complementary PSD $\tilde{\mathbf{R}}_X(f)$ is defined as a matrix-valued function of $f \in \mathcal{F}$, whose (k, l) th entry is given by $[\tilde{\mathbf{R}}_X(f)]_{k,l} \triangleq \tilde{R}_X^{(k-l)}(f + (l - L - 1)/T)$ for $k, l = 1, \dots, 2L + 1$.

Note that the matrix-valued complementary PSD $\tilde{\mathbf{R}}_N(f)$ of the interference-plus-noise signal $N(t)$ becomes an all-zero matrix because $N(t)$ is modeled by a zero-mean proper-complex SOCS random process. Note also that the effective matrix-valued complementary PSD can be defined similarly to the effective matrix-valued PSD. In what follows, each of the VFT, the matrix-valued PSD, and the matrix-valued complementary PSD is an effective one.

By using the above definitions, the matrix-valued PSD and the matrix-valued complementary PSD of the desired signal component in (1) are derived as follows.

Lemma 1: The $\mathcal{N}(f)$ -by- $\mathcal{N}(f)$ matrix-valued PSD $\mathbf{R}_X(f)$ and the $\mathcal{N}(f)$ -by- $\mathcal{N}(-f)$ matrix-valued complementary PSD $\tilde{\mathbf{R}}_X(f)$ of the desired signal $X(t) = \sum_{l=-\infty}^{\infty} b[l]p(t - lT)$ are given by

$$\mathbf{R}_X(f) = \frac{1}{T} M(fT) \mathbf{p}(f) \mathbf{p}(f)^{\mathcal{H}} \quad \text{and} \quad \tilde{\mathbf{R}}_X(f) = \frac{1}{T} \tilde{M}(fT) \mathbf{p}(f) (\mathbf{J}(-f) \mathbf{p}(-f)^*)^{\mathcal{H}}, \quad (7)$$

respectively, where $\mathbf{p}(f)$ denotes the VFT of $p(t)$, $\mathbf{J}(f)$ denotes the $\mathcal{N}(f)$ -by- $\mathcal{N}(f)$ backward identity matrix whose (m, n) th entry is given by 1 for $m + n = \mathcal{N}(f) + 1$, and 0 otherwise, and $^{\mathcal{H}}$ denotes

Hermitian transposition.

Proof: By using the CTFT of $r_X^{(k)}(\tau)$ and $\tilde{r}_X^{(k)}(\tau)$, it can be easily shown that $R_X^{(k)}(f) = M(fT)P(f + k/T)P(f)^*/T$ and $\tilde{R}_X^{(k)}(f) = \tilde{M}(fT)P(f + k/T)P(-f)/T$. Therefore, the conclusion follows from the definitions reviewed in Section III-A and Definition 1. \square

Note that $\mathbf{J}(-f)\mathbf{p}(-f)^*$ in (7) is nothing but the VFT of $p(t)^*$. Thus, $\tilde{\mathbf{R}}_X(f)$ can be interpreted as the correlation between the frequency components at f of $X(t)$ and $X(t)^*$.

Now, we are ready to convert the objective function. The MSE $\varepsilon \triangleq \varepsilon(s(t), w_1(t), w_2(t))$ defined in (4) can be rewritten as

$$\varepsilon = \mathbb{E}\{|b[l]|^2\} - 2\Re\left(\mathbb{E}\{b[l]^* z_1[l]\}\right) + \mathbb{E}\{|z_1[l]|^2\} - 2\Re\left(\mathbb{E}\{b[l]^* z_2[l]\}\right) + 2\Re\left(\mathbb{E}\{z_1[l] z_2[l]^*\}\right) + \mathbb{E}\{|z_2[l]|^2\},$$

where $\Re(\cdot)$ denotes the real part. In the following propositions, each component of the right side of (8) is expressed in terms of the VFT, the matrix-valued PSD, and the matrix-valued complementary PSD. \square

Proposition 1: The first three terms of the right side of (8) can be rewritten as $\mathbb{E}\{|b[l]|^2\} = \int_{\mathcal{F}} TM(fT)df$, $\mathbb{E}\{b[l]^* z_1[l]\} = \int_{\mathcal{F}} \mathbf{w}_1(f)^H M(fT)\mathbf{p}(f)df$, and $\mathbb{E}\{|z_1[l]|^2\} = \int_{\mathcal{F}} \mathbf{w}_1(f)^H \mathbf{R}(f)\mathbf{w}_1(f)df$, respectively, where $\mathbf{w}_1(f)$ is the VFT of $w_1(t)$ and $\mathbf{R}(f) \triangleq \mathbf{R}_N(f) + \mathbf{R}_X(f)$.

Proof: See [8, Proposition 1-4]. \square

Proposition 2: The last three terms of the right side of (8) can be rewritten as $\mathbb{E}\{b[l]^* z_2[l]\} = \int_{\mathcal{F}} \mathbf{w}_2(f)^H \tilde{M}(fT)^* \mathbf{J}(-f)\mathbf{p}(-f)^* df$, $\mathbb{E}\{z_1[l] z_2[l]^*\} = \int_{\mathcal{F}} \mathbf{w}_1(f)^H \tilde{\mathbf{R}}(f)\mathbf{w}_2(f)df$, and $\mathbb{E}\{|z_2[l]|^2\} = \int_{\mathcal{F}} \mathbf{w}_2(f)^H \mathbf{J}(-f) \mathbf{R}(-f)^* \mathbf{J}(-f)\mathbf{w}_2(f)df$, respectively, where $\mathbf{w}_2(f)$ is the VFT of $w_2(t)$ and $\tilde{\mathbf{R}}(f) \triangleq \tilde{\mathbf{R}}_X(f)$.

Proof: It can be shown similarly to Proposition 1. \square

Note in $\mathbb{E}\{|z_2[l]|^2\} = \int_{\mathcal{F}} \mathbf{w}_2(f)^H \mathbf{J}(-f) \mathbf{R}(-f)^* \mathbf{J}(-f)\mathbf{w}_2(f)df$ that the pre-multiplication of the backward identity matrix $\mathbf{J}(f)$ reverses the order of the rows whereas the post-multiplication reverses that of the columns. Note also that $\mathbf{p}(f) = \mathbf{H}(f)\mathbf{s}(f)$, $\forall f \in \mathcal{F}$, where $\mathbf{s}(f)$ is the VFT of the transmit waveform $s(t)$ and $\mathbf{H}(f)$ is defined as $\mathbf{H}(f) \triangleq \text{diag}\{\mathbf{h}(f)\}$ with $\mathbf{h}(f)$ representing the VFT of $h(t)$.

To simplify the expression of the objective function, we define $\bar{\mathcal{N}}(f) \triangleq \mathcal{N}(f) + \mathcal{N}(-f)$,

$$\bar{\mathbf{s}}(f) \triangleq [\mathbf{s}(f)^T, (\mathbf{J}(-f)\mathbf{s}(-f)^*)^T]^T \quad \text{and} \quad \bar{\mathbf{w}}(f) \triangleq [\mathbf{w}_1(f)^T, \mathbf{w}_2(f)^T]^T, \quad (9)$$

where T denotes transposition. Here, the length- $\bar{\mathcal{N}}(f)$ vector-valued functions $\bar{\mathbf{s}}(f)$ and $\bar{\mathbf{w}}(f)$ are the VFT of the transmit waveform augmented by the VFT of its complex conjugate and the VFT of a

receive waveform augmented by the VFT of the other receive waveform, respectively. Also, let the $\mathcal{N}(f)$ -by- $\mathcal{N}(f)$ matrices $\bar{\mathbf{H}}(f)$, $\bar{\mathbf{M}}(f)$, and $\bar{\mathbf{R}}(f)$ be defined, respectively, as $\bar{\mathbf{H}}(f) \triangleq \text{diag}\{\mathbf{H}(f), \mathbf{J}(-f)\mathbf{H}(-f)^*\mathbf{J}(-f)\}$, $\bar{\mathbf{M}}(f) \triangleq \text{diag}\{M(f)\mathbf{I}(f), \tilde{M}(f)^*\mathbf{I}(-f)\}$, and

$$\bar{\mathbf{R}}(f) \triangleq \begin{bmatrix} \mathbf{R}(f) & \tilde{\mathbf{R}}(f) \\ \tilde{\mathbf{R}}(f)^{\mathcal{H}} & \mathbf{J}(-f)\mathbf{R}(-f)^*\mathbf{J}(-f) \end{bmatrix} \quad (10)$$

with $\mathbf{I}(f)$ denoting the $\mathcal{N}(f)$ -by- $\mathcal{N}(f)$ identity matrix and $\text{diag}\{\mathbf{A}, \mathbf{B}\}$ denoting the block diagonal matrix whose diagonal blocks are the matrices \mathbf{A} and \mathbf{B} . These notions enable us to derive the optimal receive waveforms in a straightforward way.

By substituting the results of Propositions 1 and 2 into (8), we can rewrite the objective function ε as

$$\varepsilon(\bar{\mathbf{s}}(f), \bar{\mathbf{w}}(f)) = \int_{\mathcal{F}} \left(TM(fT) + \bar{\mathbf{w}}(f)^{\mathcal{H}} \bar{\mathbf{R}}(f) \bar{\mathbf{w}}(f) - 2\Re\{\bar{\mathbf{w}}(f)^{\mathcal{H}} \bar{\mathbf{H}}(f) \bar{\mathbf{M}}(fT) \bar{\mathbf{s}}(f)\} \right) df, \quad (11)$$

which is a function of $\bar{\mathbf{s}}(f)$ and $\bar{\mathbf{w}}(f)$. Also, by using [8, Eq. (32)] and the definition of $\bar{\mathbf{s}}(f)$, we can rewrite the average transmit power \bar{P} defined in (5) as

$$\bar{P} = \frac{1}{T} \int_{\mathcal{F}} M(fT) \mathbf{s}(f)^{\mathcal{H}} \mathbf{s}(f) df = \frac{1}{2T} \int_{\mathcal{F}} M(fT) \bar{\mathbf{s}}(f)^{\mathcal{H}} \bar{\mathbf{s}}(f) df. \quad (12)$$

This leads to the equivalent joint optimization problem to find $\bar{\mathbf{s}}(f)$ and $\bar{\mathbf{w}}(f)$ as

$$\begin{aligned} & \text{minimize}_{\bar{\mathbf{s}}(f), \bar{\mathbf{w}}(f)} \varepsilon(\bar{\mathbf{s}}(f), \bar{\mathbf{w}}(f)) \\ & \text{subject to } \bar{P} = P_T. \end{aligned} \quad (13a) \quad (13b)$$

Problem 2:

In the next section, we solve this optimization problem to obtain the VFTs of the optimal receive and transmit waveforms.

IV. OPTIMIZATION OF TRANSMIT AND RECEIVE WAVEFORMS

In this section, we first derive the optimal $\bar{\mathbf{w}}(f)$ that minimizes the objective function in (13a) for a given $\bar{\mathbf{s}}(f)$. Then, by substituting this $\bar{\mathbf{w}}(f)$ and introducing the notion of the propriety frequency function, we obtain the optimization problem over $\mathbf{s}(f)$. By solving this problem, we finally obtain the optimal transmit and receive waveforms.

A. Optimization of Widely Linear Receiver

As in [8, Theorem 2], to find the optimal $\bar{\mathbf{w}}(f)$ for given $\bar{\mathbf{s}}(f)$, an unconstrained quadratic optimization problem is solved. Thus, by using the first-order necessary condition, we have the solution

$$\bar{\mathbf{w}}(f) = \bar{\mathbf{R}}(f)^{-1} \bar{\mathbf{H}}(f) \bar{\mathbf{M}}(fT) \bar{\mathbf{s}}(f), \quad \forall f \in \mathcal{F}. \quad (14)$$

By substituting the above solution into (11), we can rewrite the MSE as

$$\varepsilon(\bar{\mathbf{s}}(f)) = \int_{\mathcal{F}} \left(T M(fT) - \bar{\mathbf{s}}(f)^H \bar{\mathbf{M}}(fT)^H \bar{\mathbf{H}}(f)^H \bar{\mathbf{R}}(f)^{-1} \bar{\mathbf{H}}(f) \bar{\mathbf{M}}(fT) \bar{\mathbf{s}}(f) \right) df. \quad (15)$$

which is a function only of $\bar{\mathbf{s}}(f)$.

B. Improperity Frequency Function

To convert $\varepsilon(\bar{\mathbf{s}}(f))$ into a function only of $\mathbf{s}(f)$, the notion of the improperity frequency function is introduced as follows.

Definition 2: Given a discrete-time improper-complex SOS random process with PSD $M(f)$ and complementary PSD $\tilde{M}(f)$, its improperity frequency function $k(f)$ is defined as

$$k(f) \triangleq \begin{cases} 0, & \text{if } M(f)M(-f) = 0, \\ \frac{|\tilde{M}(f)|}{\sqrt{M(f)M(-f)}}, & \text{otherwise.} \end{cases} \quad (16)$$

The above definition is motivated by the improperity coefficient of an improper-complex random variable [13, Definition 3.1] and by a relation between $M(f)$ and $\tilde{M}(f)$ shown in [17, Eq. (5)]. By using the phase $\phi(f)$ of $\tilde{M}(f)$, we can rewrite the complementary PSD as $\tilde{M}(f) = |\tilde{M}(f)|e^{j\phi(f)} = k(f)\sqrt{M(f)M(-f)}e^{j\phi(f)}$, where $0 \leq \phi(f) \leq 2\pi$. In the next lemma, the properties of the improperity frequency and the phase functions are provided.

Lemma 2: The improperity frequency function $k(f)$ and the phase function $\phi(f)$ satisfy

$$0 \leq k(f) \leq 1, \quad k(-f) = k(f), \quad \text{and} \quad \phi(-f) = \phi(f), \quad \forall f. \quad (17)$$

Proof: Since $\tilde{m}[-k] = \tilde{m}[k]$ by definition, we have $\tilde{M}(-f) = \tilde{M}(f)$, which implies $\phi(-f) = \phi(f)$. This also leads to $k(-f) = k(f)$ by (16). By using the property $|\tilde{M}(f)|^2 \leq M(f)M(-f)$ shown in [17, Eq. (5)], we have $0 \leq k(f) \leq 1$. \square

For example, an uncorrelated real-valued PAM data sequence results in $k(f) = 1, \forall f$, whereas any proper-complex data sequence results in $k(f) = 0, \forall f$. By using the improperity frequency function, we can rewrite the MSE (15) in the form of a function of $\bar{\mathbf{s}}(f)$ as a function of $\mathbf{s}(f)$.

Lemma 3: Define $c(f)$ as

$$c(f) \triangleq \frac{1}{T} M(fT) \mathbf{s}(f)^H \mathbf{H}(f)^H \mathbf{R}_N(f)^{-1} \mathbf{H}(f) \mathbf{s}(f). \quad (18)$$

By using $c(f)$ and $k(f)$, also define $\mathbf{D}(f)$ and $\mathbf{k}(f)$, respectively, as

$$\mathbf{D}(f) \triangleq \begin{bmatrix} c(f) & 0 \\ 0 & \frac{c(-f)}{1+c(-f)(1-k(fT)^2)} \end{bmatrix} \quad \text{and} \quad \mathbf{k}(f) \triangleq \begin{bmatrix} 1 \\ k(f) \end{bmatrix}. \quad (19)$$

Then, the MSE $\varepsilon(\bar{\mathbf{s}}(f))$ in (15) can be rewritten as

$$\varepsilon(\mathbf{s}(f)) = \int_{\mathcal{F}} \frac{TM(fT)}{1 + \mathbf{k}(fT)^T \mathbf{D}(f) \mathbf{k}(fT)} df, \quad (20)$$

which is a function of $\mathbf{s}(f)$.

Proof: See Appendix B. □

C. Optimization of Transmitter

Let $\varepsilon(f)$ denote the integrand in (20), i.e., $\varepsilon(f) \triangleq TM(fT)/(1 + \mathbf{k}(fT)^T \mathbf{D}(f) \mathbf{k}(fT))$. Then, by the definitions of $c(f)$ and $\mathbf{D}(f)$ in (18) and (19), respectively, it can be seen that $\varepsilon(f_0)$ for some f_0 is affected by the choice of $\mathbf{s}(f)$ at both f_0 and $-f_0$. Thus, Problem 2 can be rewritten as

$$\text{Problem 3:} \quad \begin{aligned} & \underset{a(f), a(-f)}{\text{minimize}} \quad \int_{\mathcal{F}^+} \varepsilon(f) + \varepsilon(-f) df \\ & \text{subject to} \quad \|\mathbf{s}(f)\|^2 = a(f), \quad \forall f \in \mathcal{F} \end{aligned} \quad (21a)$$

$$\text{subject to} \quad \frac{1}{T} \int_{\mathcal{F}^+} M(fT)a(f) + M(-fT)a(-f)df = P_T, \quad (21b)$$

where $a(f) \triangleq \|\mathbf{s}(f)\|^2$ is the energy density of $\mathbf{s}(f)$ and $\mathcal{F}^+ \triangleq \{f : 0 \leq f < 1/(2T)\}$ denotes the half-Nyquist interval. Note that the problem is now in the form of a double minimization problem, where the constraint set of $\mathbf{s}(f)$ is partitioned into subsets, each of which has all $\mathbf{s}(f)$ having the same $a(f)$.

Proposition 3: Given $a(f)$, the optimal solution to the inner optimization problem in (21a) is given by

$$\mathbf{s}(f) = \sqrt{a(f)} \mathbf{v}(f) e^{j\theta(f)}, \quad \forall f \in \mathcal{F}, \quad (22)$$

where $\mathbf{v}(f)$ is the normalized eigenvector corresponding to the largest eigenvalue of $\mathbf{H}(f)^H \mathbf{R}_N(f)^{-1} \mathbf{H}(f)$, and $\theta(f)$ can be chosen arbitrarily.

Proof: Note that the integrand $\varepsilon(f) + \varepsilon(-f)$ in (21a) can be rewritten as

$$\varepsilon(f) + \varepsilon(-f) = T \frac{M(-fT)(1 + c(f)\bar{k}(fT)) + M(fT)(1 + c(-f)\bar{k}(fT))}{1 + c(f) + c(-f) + c(f)c(-f)\bar{k}(fT)}, \quad (23)$$

where $\bar{k}(f) \triangleq 1 - k(f)^2$. Since $\varepsilon(f) + \varepsilon(-f)$ evaluated at some f_0 is a function only of $\mathbf{s}(f_0)$ and $\mathbf{s}(-f_0)$ through $c(f_0)$ and $c(-f_0)$, respectively, we just need to minimize by optimizing $c(f_0)$ and $c(-f_0)$ in the integrand at each $f \in \mathcal{F}^+$ subject to the constraint. Let $c(f_0) = c_1$ and $c(-f_0) = c_2$. Then, it can be shown that $\partial(\varepsilon(f_0) + \varepsilon(-f_0))/\partial c_1 < 0$ and $\partial(\varepsilon(f_0) + \varepsilon(-f_0))/\partial c_2 < 0$. Moreover, since $a(f_0) = \mathbf{s}(f_0)^H \mathbf{s}(f_0)$, $c(f_0)$ is constrained by $a(f_0)$ through $\mathbf{s}(f_0)$ and $c(-f_0)$ is constrained by $a(-f_0)$ through $\mathbf{s}(-f_0)$. Thus, we now can separately find $\mathbf{s}(f_0)$ that maximizes $c(f_0)$ for given $a(f_0)$ and $\mathbf{s}(-f_0)$ that maximizes $c(-f_0)$

for given $a(-f_0)$. This maximization of $c(f)$ defined in (18) subject to $a(f) = \|\mathbf{s}(f)\|^2$ is exactly the same problem solved in [8, Section IV-B], where the optimal solution is given by (22) at each $f \in \mathcal{F}$. Therefore, the conclusion follows. \square

According to (22), the optimal $\mathbf{s}(f)$ given $a(f)$ is not affected by the improper frequency function $k(f)$. However, it actually affects the outer optimization of $a(f)$, which will be performed in what follows.

Let $\lambda(f)$ denote the largest eigenvalue of $\mathbf{H}(f)^H \mathbf{R}_N(f)^{-1} \mathbf{H}(f)$. Then, by (22), $c(f)$ can be simplified

as $c(f) = M(fT)\lambda(f)a(f)/T, \forall f \in \mathcal{F}$. Thus, the outer minimization problem of Problem 3 to find the optimal energy density $a_{\text{opt}}(f)$ for $\int_{\mathcal{F}} \bar{\varepsilon}(f) df$ becomes (24a)

$$\text{Problem 4:} \quad \text{subject to } \frac{1}{T} \int_{\mathcal{F}^+} M(fT)a(f) + M(-fT)a(-f)df = P_T, \quad (24b)$$

where $\bar{\varepsilon}(f)$ is given by

$$\bar{\varepsilon}(f) \triangleq \frac{\frac{M(-fT)}{T} \left(1 + \frac{M(fT)}{T} \lambda(f)a(f)\bar{k}(fT)\right) + \frac{M(fT)}{T} \left(1 + \frac{M(-fT)}{T} \lambda(-f)a(-f)\bar{k}(fT)\right)}{1 + \frac{M(fT)}{T} \lambda(f)a(f) + \frac{M(-fT)}{T} \lambda(-f)a(-f) + \frac{M(fT)}{T} \lambda(f)a(f) \frac{M(-fT)}{T} \lambda(-f)a(-f)\bar{k}(fT)}, \quad (25)$$

with $\bar{k}(f) \triangleq 1 - k(f)^2$ as already used in (23). Now, we are ready to present the optimal $a(f)$. In what follows, \mathcal{F}_M and \mathcal{F}_λ denote the supports of $M(fT)$ and $\lambda(f)$, respectively, i.e., $\mathcal{F}_M \triangleq \{f \in \mathcal{F} : M(fT) \neq 0\}$ and $\mathcal{F}_\lambda \triangleq \{f \in \mathcal{F} : \lambda(f) \neq 0\}$.

Proposition 4: The optimal solution to Problem 4 can be found by performing a line search for a parameter ν in $(0, \nu_{\max}]$, where $\nu_{\max} \triangleq \max_f \lambda(f)(M(-fT)k(fT)^2 + M(fT))$. For each $\nu \in (0, \nu_{\max}]$, a candidate density function can be constructed by using the algorithm described in Table I, where

$$u(\nu, f) = \begin{cases} \left[\frac{1}{\sqrt{\nu}} - \frac{1 + M(-fT)\lambda(-f)a(-f)}{\sqrt{\lambda(f)\nu(f)}} \right]^+ \frac{\sqrt{\lambda(f)\nu(f)}}{\lambda(f)M(fT)g(-f)}, & \text{for } f \in \mathcal{F}_M \cap \mathcal{F}_\lambda, \\ 0, & \text{for } f \in \mathcal{F}_M \cap (\mathcal{F}_\lambda)^c, \\ \text{arbitrary,} & \text{for } f \in (\mathcal{F}_M)^c, \end{cases} \quad (26)$$

with $g(f) \triangleq 1 + M(fT)\lambda(f)a(f)\bar{k}(fT)$, $\nu(f) \triangleq M(-fT)k(fT)^2 + M(fT)g(-f)^2$, and $[x]^+ \triangleq \max(x, 0)$.

The candidate function that satisfies the power constraint (24b) is the optimal density function $a_{\text{opt}}(f)$.

Proof: See Appendix C. \square

Note that any line search algorithm can be used to find $a_{\text{opt}}(f)$ in Proposition 4. Note also that the algorithm in Table I allows the construction of an approximate solution with arbitrary accuracy if the interval \mathcal{F}^+ is partitioned finely enough.

Now, by using $a_{\text{opt}}(f)$, we can find the VFTs of the optimal transmit and receive waveforms as follows.

Theorem 1: The VFT $\mathbf{s}_{\text{opt}}(f)$ of the jointly optimal transmit waveform $s_{\text{opt}}(t)$ as the solution to Problem 1 are given by

$$\mathbf{s}_{\text{opt}}(f) = \begin{cases} \sqrt{a_{\text{opt}}(f)} \mathbf{v}(f) e^{j\theta(f)}, & \text{for } f \in \mathcal{F}_M, \\ \text{arbitrary}, & \text{for } f \in (\mathcal{F}_M)^c, \end{cases} \quad (27)$$

where $\theta(f)$ can be chosen arbitrarily. Then, the VFTs $\mathbf{w}_{1,\text{opt}}(f)$ and $\mathbf{w}_{2,\text{opt}}(f)$ of the jointly optimal receive waveforms $w_{1,\text{opt}}(t)$ and $w_{2,\text{opt}}(t)$ can be found by using (14).

Proof: The conclusion immediately follows from the relation (9) among $\mathbf{s}_{\text{opt}}(f)$, $\bar{\mathbf{s}}_{\text{opt}}(f)$, $\mathbf{w}_{1,\text{opt}}(f)$, $\mathbf{w}_{2,\text{opt}}(f)$, and $\bar{\mathbf{w}}_{\text{opt}}(f)$, and Propositions 3 and 4. \square

As already mentioned, cyclostationarity and impropriety, respectively, imply the periodic spectral correlation and the symmetric spectral correlation about the origin [13, Ch. 10], [18]. Theorem 1 vividly shows these structures in the optimal transmitted signal. Specifically, the use of the VFT technique and the augmentation of $s(f)$ and $s(-f)$ to form $\bar{s}(f)$ take care of the periodic spectral correlation and the symmetric spectral correlation, respectively.

V. NUMERICAL RESULTS

In this section, numerical results are provided that show the magnitude square of the optimal transmit and receive waveforms and that show the MSE performance achieved by the optimal waveforms as a function of the amount of impropriety. For illustrative purposes, it is assumed throughout this section that an interferer linearly modulates a data sequence consisting of uncorrelated zero-mean proper-complex QPSK symbols with $E_s/N_0 = 10$ [dB] and a square-root raised cosine transmit waveform having excess bandwidth $\beta = 0.25$. It is assumed that the Tx linearly modulates a data sequence consisting of uncorrelated zero-mean improper-complex QAM symbols with uncorrelated in-phase and quadrature components. It is also assumed that all the channels are frequency flat and corrupted by AWGN.

The first results are to compare the PSD of data-like interference with the squared magnitudes of the optimal transmit and receive waveforms. There is a single interferer in Figs. 2-(a) and (b), whereas there are two uncorrelated interferers in Fig. 2-(c). The QAM symbols of the Tx have $E_s/N_0 = 5$ [dB]. For Fig. 2-(a), the QAM symbols have the in-phase variance the same as the quadrature variance, which implies $k(f) = 0, \forall f$. For Figs. 2-(b) and (c), the QAM symbols have the in-phase variance 4-times the

quadrature variance, which implies $k(f) = 0.8, \forall f$. It can be seen that $w_{2,\text{opt}}(t)$, processing the complex conjugate of the received signal, is zero for the data sequence having $k(f) = 0, \forall f$, but it is non-zero for the data sequence having $k(f) = 0.8, \forall f$.

The next results are to compare the MSEs achieved by the optimal transmit and receive waveforms for different levels of impropriety. We consider the same number of interferers and interference parameters as Fig. 2-(c). In Fig. 3-(a), the QAM symbols of the Tx have E_s/N_0 from 0 to 15 [dB] and have $k(f) = 0.0, 0.2, 0.4, 0.6, 0.8$, or $1.0, \forall f$. In Fig. 3-(b), the QAM symbols of the Tx have $E_s/N_0 = 0, 5, 10$ or 15 [dB] and have $k(f)$ from 0 to 1, $\forall f$. In both cases, as the amount of impropriety increases, the optimal pair of the Tx and Rx more exploits impropriety and cyclostationarity of the desired signal in suppressing the data-like interference and, consequently, the MSE performance monotonically improves.

VI. CONCLUSIONS

In this paper, we have considered a joint optimization of the Tx and Rx for the transmission of an improper-complex SOS data sequence over an additive proper-complex cyclostationary noise channel. An MSE minimization problem is formulated under the average transmit power constraint to find the jointly optimal transmit waveform of a linear modulator and the receive waveforms of a widely linear Rx. This problem is converted into an equivalent problem described in the frequency domain with the help of the VFT technique and solved by introducing the notion of the impropriety frequency function. It is shown that the optimal transmit and receive waveforms well exploit the frequency-domain second-order structure of the improper-complex SOS data sequence and the additive proper-complex SOCS noise.

APPENDIX

A. Proof of Lemma 3

Proof: Define the 2-by-2 matrices $\hat{\mathbf{M}}(f)$, $\mathbf{M}(f)$, and $\mathbf{K}(f)$, respectively, as

$$\hat{\mathbf{M}}(f) \triangleq \begin{bmatrix} M(f) & \tilde{M}(f) \\ \tilde{M}(f)^* & M(-f)^* \end{bmatrix}, \mathbf{M}(f) \triangleq \begin{bmatrix} M(f)e^{j\phi(f)} & 0 \\ 0 & M(-f)e^{-j\phi(f)} \end{bmatrix}^{\frac{1}{2}}, \text{ and } \mathbf{K}(f) \triangleq \begin{bmatrix} 1 & k(f) \\ k(f) & 1 \end{bmatrix}. \quad (28)$$

Then, we can rewrite $\hat{\mathbf{M}}(f)$ as $\hat{\mathbf{M}}(f) = \mathbf{M}(f)\mathbf{K}(f)\mathbf{M}(f)^H$. Also, define the $\bar{\mathcal{N}}(f)$ -by- $\bar{\mathcal{N}}(f)$ matrix $\bar{\mathbf{R}}_N(f)$ and the $\bar{\mathcal{N}}(f)$ -by-2 matrix $\bar{\mathbf{S}}(f)$ as $\bar{\mathbf{R}}_N(f) \triangleq \text{diag}\{\mathbf{R}_N(f), \mathbf{J}(-f)\mathbf{R}_N(-f)^*\mathbf{J}(-f)\}$ and

$\bar{\mathbf{S}}(f) \triangleq \text{diag}\{\mathbf{s}(f), \mathbf{J}(-f)\mathbf{s}(-f)^*\}$, respectively. Due to the ambient noise component in $N(t)$, $\mathbf{R}_N(f)$ and $\bar{\mathbf{R}}_N(f)$ are positive definite for all $f \in \mathcal{F}$. By using $\bar{\mathbf{R}}_N(f)^{-1/2}$, define the $\bar{\mathcal{N}}(f)$ -by-2 matrix $\tilde{\mathbf{P}}(f)$ as $\tilde{\mathbf{P}}(f) \triangleq \bar{\mathbf{R}}_N(f)^{-1/2} \bar{\mathbf{H}}(f) \bar{\mathbf{S}}(f) \mathbf{M}(fT) / \sqrt{T}$. Then, it can be shown that $\bar{\mathbf{H}}(f) \bar{\mathbf{M}}(fT) \bar{\mathbf{s}}(f) = \sqrt{T \mathbf{M}(fT)} e^{-j\phi(fT)/2} \bar{\mathbf{R}}_N(f)^{1/2} \tilde{\mathbf{P}}(f) \mathbf{k}(fT)$. Thus, the second term of the integrand in (15), which contains $\bar{\mathbf{H}}(f) \bar{\mathbf{M}}(fT) \bar{\mathbf{s}}(f)$, can be rewritten as

$$\bar{\mathbf{s}}(f)^{\mathcal{H}} \bar{\mathbf{M}}(fT)^{\mathcal{H}} \bar{\mathbf{H}}(f)^{\mathcal{H}} \bar{\mathbf{R}}(f)^{-1} \bar{\mathbf{H}}(f) \bar{\mathbf{M}}(fT) \bar{\mathbf{s}}(f)$$

$$= TM(fT) \mathbf{k}(fT)^{\mathcal{T}} \tilde{\mathbf{P}}(f)^{\mathcal{H}} \left(\mathbf{I} + \tilde{\mathbf{P}}(f) \mathbf{K}(fT) \tilde{\mathbf{P}}(f)^{\mathcal{H}} \right)^{-1} \tilde{\mathbf{P}}(f) \mathbf{k}(fT), \quad (29)$$

where \mathbf{I} conveniently denotes the appropriately sized identity matrix throughout this proof. Let $\hat{\mathbf{p}}(f) \triangleq \sqrt{\mathbf{M}(fT)/T} e^{j\phi(f)/2} \mathbf{R}_N(f)^{-1/2} \mathbf{H}(f) \mathbf{s}(f)$. Then, we can rewrite $\tilde{\mathbf{P}}(f)$ and $c(f)$ defined in (18) as $\tilde{\mathbf{P}}(f) = \text{diag}\{\hat{\mathbf{p}}(f), \mathbf{J}(-f)\hat{\mathbf{p}}(-f)^*\}$ and $c(f) = \|\hat{\mathbf{p}}(f)\|^2$, respectively. If $c(f)c(-f) = 0$, then it can be shown that (29) leads to (20) by using the matrix inversion lemma showing $\mathbf{I} - \mathbf{u}^{\mathcal{H}}(\mathbf{I} + \mathbf{u}\mathbf{u}^{\mathcal{H}})^{-1}\mathbf{u} = (1 + \mathbf{u}^{\mathcal{H}}\mathbf{u})^{-1}$ for any vector \mathbf{u} . If $c(f)c(-f) \neq 0$, then, since $\tilde{\mathbf{P}}(f)^{\mathcal{H}} \tilde{\mathbf{P}}(f) = \text{diag}\{c(f), c(-f)\}$ is invertible, it can be shown that $\tilde{\mathbf{P}}(f)^{\mathcal{H}} (\mathbf{I} + \tilde{\mathbf{P}}(f) \mathbf{K}(fT) \tilde{\mathbf{P}}(f)^{\mathcal{H}})^{-1} \tilde{\mathbf{P}}(f) \tilde{\mathbf{C}}(f) = \mathbf{I}$, where $\tilde{\mathbf{C}}(f)$ is defined as $\tilde{\mathbf{C}}(f) \triangleq (\tilde{\mathbf{P}}(f)^{\mathcal{H}} \tilde{\mathbf{P}}(f))^{-1} + \mathbf{K}(fT)$. Since $c(f)$ and $c(-f)$ are not zero, we can rewrite $\tilde{\mathbf{C}}(f)$ as $\tilde{\mathbf{C}}(f) = \text{diag}\{c(f)^{-1}, c(-f)^{-1} + 1 - k(fT)^2\} + \mathbf{k}(fT) \mathbf{k}(fT)^{\mathcal{T}}$. Thus, we now can rewrite the right side of (29) as $TM(fT) \mathbf{k}(fT)^{\mathcal{T}} \tilde{\mathbf{C}}(f)^{-1} \mathbf{k}(fT)$. By using the matrix inversion lemma, the conclusion follows. \square

B. Proof of Proposition 4

Proof: For convenience, the integration interval \mathcal{F}^+ is partitioned into N equal-length subintervals. Then, the solution can be straightforwardly extended to the original problem by letting N tend to infinity. Let $\xi_i \triangleq i/(2NT) - 1/(4NT)$, $a_i \triangleq a(\xi_i)$, $\hat{a}_i \triangleq a(-\xi_i)$, $m_i \triangleq M(\xi_i T)/T$, $\hat{m}_i \triangleq M(-\xi_i T)/T$, $\lambda_i \triangleq \lambda(\xi_i)$, $\hat{\lambda}_i \triangleq \lambda(-\xi_i)$, and $k_i \triangleq k(\xi_i T)$. Then, the original optimization problem can be approximated by

$$\underset{a_i, \hat{a}_i \geq 0}{\text{minimize}} \quad \sum_{i=1}^N f_i(a_i, \hat{a}_i) \quad \text{subject to} \quad \sum_{i=1}^N (m_i a_i + \hat{m}_i \hat{a}_i) \leq P_T T, \quad (30)$$

where $f_i(a_i, \hat{a}_i)$ is given by $f_i(a_i, \hat{a}_i) \triangleq (\hat{m}_i(1 + m_i \lambda_i a_i \bar{k}_i) + m_i(1 + \hat{m}_i \hat{\lambda}_i \hat{a}_i \bar{k}_i)) / (1 + m_i \lambda_i a_i + \hat{m}_i \hat{\lambda}_i \hat{a}_i + m_i \lambda_i a_i \hat{m}_i \hat{\lambda}_i \hat{a}_i \bar{k}_i)$ for non-negative real numbers $m_i, \hat{m}_i, \lambda_i, \hat{\lambda}_i$, and $\bar{k}_i \triangleq 1 - k_i^2$ with $0 \leq k_i < 1, \forall i$. It can be easily shown that, if $m_i = 0$, a_i can be chosen arbitrarily because a_i does not affect both the objective function and the constraint. It can be also easily shown that $\lambda_i = 0$ results in $a_i = 0$ to keep from wasting

the transmit power. Similarly, if $\hat{m}_i = 0$, then \hat{a}_i can be chosen arbitrarily, and if $\hat{\lambda}_i = 0$, then $\hat{a}_i = 0$. The case of $k_i = 1$ is discussed after solving the optimization problem for $k_i < 1$. Thus, in what follows, we assume that $m_i \hat{m}_i \neq 0$, $\lambda_i \hat{\lambda}_i \neq 0$, and $k_i < 1$, $\forall i$.

Define \mathbf{a} and \mathbf{m} as $\mathbf{a} \triangleq [a_1, \hat{a}_1, a_2, \hat{a}_2, \dots, a_N, \hat{a}_N]^T$ and $\mathbf{m} \triangleq [m_1, \hat{m}_1, m_2, \hat{m}_2, \dots, m_N, \hat{m}_N]^T$, respectively. Then, it can be easily shown that the Hessian $\mathbf{F}_i(\mathbf{a})$ of the objective function $\sum_{i=1}^N f_i(a_i, \hat{a}_i)$ is a positive definite matrix for each \mathbf{a} and the equality constraint $\sum_{i=1}^N (m_i a_i + \hat{m}_i \hat{a}_i) = \mathbf{m}^T \mathbf{a}$ is an affine function of \mathbf{a} . Thus, the problem in (30) is a strictly convex optimization problem. Since the Karush-Kuhn-Tucker (KKT) condition is necessary and sufficient for a point to be the unique solution of a strictly convex optimization problem [19, Theorem 22.9], we first need to find the KKT condition.

The Lagrangian function of (30) can be written as $l(\mathbf{a}, \nu, \boldsymbol{\mu}) = \sum_{i=1}^N f_i(a_i, \hat{a}_i) + \nu(\mathbf{m}^T \mathbf{a} - P_T T) - \boldsymbol{\mu}^T \mathbf{a}$ by introducing the multipliers ν and $\boldsymbol{\mu} \triangleq [\mu_1, \hat{\mu}_1, \mu_2, \hat{\mu}_2, \dots, \mu_N, \hat{\mu}_N]^T$. Then, the KKT condition can be written as $-m_i \lambda_i (\hat{m}_i k^2 + m_i \hat{g}_i(\hat{a}_i)^2) / h_i(a_i, \hat{a}_i)^2 + m_i \nu - \mu_i = 0$ and $-\hat{m}_i \hat{\lambda}_i (m_i k^2 + \hat{m}_i g_i(a_i)^2) / h_i(a_i, \hat{a}_i)^2 + \hat{m}_i \nu - \hat{\mu}_i = 0$ with $a_i \geq 0$, $\hat{a}_i \geq 0$, $\mu_i \geq 0$, $\hat{\mu}_i \geq 0$, $\mu_i a_i = 0$, $\hat{\mu}_i \hat{a}_i = 0$, $\forall i$, and $\sum_{i=1}^N (m_i a_i + \hat{m}_i \hat{a}_i) = P_T T$, where $h_i(a_i, \hat{a}_i) \triangleq 1 + m_i \lambda_i a_i + \hat{m}_i \hat{\lambda}_i \hat{a}_i + m_i \lambda_i a_i \hat{m}_i \hat{\lambda}_i \hat{a}_i \bar{k}_i$, $g_i(a_i) \triangleq 1 + m_i \lambda_i a_i \bar{k}_i$, and $\hat{g}_i(\hat{a}_i) \triangleq 1 + \hat{m}_i \hat{\lambda}_i \hat{a}_i \bar{k}_i$.

Define $\nu_i(a_i, \hat{a}_i) \triangleq \lambda_i (\hat{m}_i k^2 + m_i \hat{g}_i(\hat{a}_i)^2) / h_i(a_i, \hat{a}_i)^2$ and $\hat{\nu}_i(a_i, \hat{a}_i) \triangleq \hat{\lambda}_i (m_i k^2 + \hat{m}_i g_i(a_i)^2) / h_i(a_i, \hat{a}_i)^2$, respectively. It can be easily shown that $\partial \nu_i(a_i, \hat{a}_i) / \partial a_i < 0$, $\partial \nu_i(0, \hat{a}_i) / \partial \hat{a}_i < 0$, $\partial \hat{\nu}_i(a_i, \hat{a}_i) / \partial \hat{a}_i < 0$, and $\partial \hat{\nu}_i(a_i, 0) / \partial a_i < 0$ for all $a_i \geq 0$ and $\hat{a}_i \geq 0$. Thus, $\nu_i(a_i, \hat{a}_i) < \nu_i(0, 0)$ and $\hat{\nu}_i(a_i, \hat{a}_i) < \hat{\nu}_i(0, 0)$, respectively, for all $a_i > 0$ and $\hat{a}_i > 0$. It can be also shown that, if $\nu \geq \nu_i(0, 0)$ and $\nu \geq \hat{\nu}_i(0, 0)$, then only $a_i = 0$ and $\hat{a}_i = 0$ satisfy the KKT condition. It is noteworthy that ν satisfying the KKT condition is upper-bounded by ν_{\max} that is defined as the largest value among $\nu_i(0, 0)$ and $\hat{\nu}_i(0, 0)$, $\forall i$, which can be easily found and is finite and positive. Thus, to find \mathbf{a} , ν , and $\boldsymbol{\mu}$ that jointly satisfy the KKT condition, a line search for ν can be performed over the interval $(0, \nu_{\max}]$, where two steps are needed to construct a candidate solution \mathbf{a} and the multiplier $\boldsymbol{\mu}$ at each ν .

First, a candidate solution \mathbf{a} associated with ν is constructed as follows. Given ν , we need to find the pair of (a_i, \hat{a}_i) satisfying the KKT condition, i.e., $\nu - \mu_i / m_i = \nu_i(a_i, \hat{a}_i)$ and $\nu - \hat{\mu}_i / \hat{m}_i = \hat{\nu}_i(a_i, \hat{a}_i)$ with $a_i \geq 0$, $\hat{a}_i \geq 0$, $\mu_i \geq 0$, $\hat{\mu}_i \geq 0$, $\mu_i a_i = 0$, $\hat{\mu}_i \hat{a}_i = 0$, which can be rewritten as

$$a_i = u_1(\hat{a}_i) \triangleq \left[\sqrt{\lambda_i(\hat{m}_i k^2 + m_i \hat{g}_i(\hat{a}_i)^2)} / \sqrt{\nu} - (1 + \hat{m}_i \hat{\lambda}_i \hat{a}_i) \right]^+ (\lambda_i m_i \hat{g}_i(\hat{a}_i))^{-1}, \quad (31a)$$

$$\hat{a}_i = u_2(a_i) \triangleq \left[\sqrt{\hat{\lambda}_i(m_i k^2 + \hat{m}_i g(a_i)^2)} / \sqrt{\nu} - (1 + m_i \lambda_i a_i) \right]^+ (\hat{\lambda}_i \hat{m}_i g(a_i))^{-1}, \quad (31b)$$

$\mu_i = 0$ if $a_i > 0$, and $\hat{\mu}_i = 0$ if $\hat{a}_i > 0$. It can be easily shown that $u_1(\hat{a}_i)$ is a decreasing function of \hat{a}_i and $u_2(a_i)$ is a decreasing function of a_i . Thus, $(u_2 \circ u_1)(\hat{a}_i)$ becomes an increasing function of \hat{a}_i . It is noteworthy that the non-negative numbers a_i and \hat{a}_i are upper-bounded by $u_1(0)$ and $u_2(0)$. Thus, when we alternately update a_i and \hat{a}_i from $\hat{a}_i = 0$ by using (31a) and (31b), respectively, both a_i and \hat{a}_i converge to the solution satisfying the KKT conditions. This iteration algorithm can be also used to find the candidate solution a_i and \hat{a}_i for the case of $k_i = 1$. Note that, if $k_i = 1$ and $\lambda_i = \hat{\lambda}_i$, any pair of a_i and \hat{a}_i satisfying $m_i a_i + \hat{m}_i \hat{a}_i = [\sqrt{(m_i + \hat{m}_i)/(\lambda_i \nu)} - 1/\lambda_i]^+$ can be the candidate solution associated with ν . After finding a_i and \hat{a}_i , μ_i and $\hat{\mu}_i$ can be computed by substituting a_i , \hat{a}_i , and ν into the KKT condition.

Second, after constructing the candidate solution \mathbf{a} associated with ν , we check whether the candidate solution satisfies the power constraint $\sum_{i=1}^N (m_i a_i + \hat{m}_i \hat{a}_i) = P_T T$. If so, then the candidate solution associated with ν_{opt} is the optimal solution \mathbf{a}_{opt} . If not, then the line search continues. Therefore, the conclusion follows. \square

REFERENCES

- [1] J. G. Proakis, *Digital Communications*, 4th ed. NY: McGraw Hill, 2001.
- [2] E. Serpedin, F. Panduru, I. Sari, and G. B. Giannakis, "Bibliography on cyclostationarity," *Signal Process.*, vol. 85, no. 12, pp. 2233-2303, Dec. 2005.
- [3] W. A. Gardner, A. Napolitano, and L. Paura, "Cyclostationarity: Half a century of research," *Signal Process.*, vol. 86, no. 4, pp. 639-697, Apr. 2006.
- [4] T. Berger and D. W. Tufts, "Optimum pulse amplitude modulation Part I: Transmitter-receiver design and bounds from information theory," *IEEE Trans. Inf. Theory*, vol. 13, no. 2, pp. 196-208, Apr. 1967.
- [5] E. Hänsler, "Some properties of transmission systems with minimum mean-square error," *IEEE Trans. Commun. Technol.*, vol. 19, no. 4, pp. 576-579, Aug. 1971.
- [6] T. Ericson, "Optimum PAM filters are always band limited," *IEEE Trans. Inf. Theory*, vol. 19, no. 4, pp. 570-573, July 1973.
- [7] T. Ericson and U. Johansson, "A general time-discrete equivalent to a time-continuous Gaussian channel," *IEEE Trans. Inf. Theory*, vol. 20, no. 4, pp. 544-549, July 1974.

- [8] J. H. Cho, "Joint transmitter and receiver optimization in additive cyclostationary noise," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3396-3405, Dec. 2004.
- [9] W. A. Gardner and L. E. Franks, "Characterization of cyclostationary random signal processes," *IEEE Trans. Inf. Theory*, vol. 21, no. 1, pp. 4-14, Jan. 1975.
- [10] J. Yang and S. Roy, "On joint transmitter and receiver optimization for multiple-input-multiple-output (MIMO) transmission systems," *IEEE Trans. Commun.*, vol. 42, no. 12, pp. 3221-3231, Dec. 1994.
- [11] G. D. Golden, J. E. Mazo, and J. Salz, "Transmitter design for data transmission in the presence of a data-like interferer," *IEEE Trans. Commun.*, vol. 43, no. 2-4, pp. 837-850, Feb.-Apr. 1995.
- [12] F. D. Neeser and J. L. Massey, "Proper complex random processes with applications to information theory," *IEEE Trans. Inf. Theory*, vol. 39, no. 4, pp. 1293-1302, July 1993.
- [13] P. J. Schreier and L. L. Scharf, *Statistical Signal Processing of Complex-Valued Data: The Theory of Improper and Noncircular Signals*. NY: Cambridge Univ. Press, 2010.
- [14] W. M. Brown and R. B. Crane, "Conjugate linear filtering," *IEEE Trans. Inf. Theory*, vol. 15, no. 4, pp. 462-465, July 1969.
- [15] Y. H. Yun and J. H. Cho, "An optimal orthogonal overlay for a cyclostationary legacy signal," *IEEE Trans. Commun.*, vol. 58, no. 5, pp. 1557-1567, May 2010.
- [16] J. H. Cho and W. Gao, "Continuous-time equivalents of Welch bound equality sequences," *IEEE Trans. Inf. Theory*, vol. 51, no. 9, pp. 3176- 3185, Sept. 2005.
- [17] B. Picinbono and P. Bondon, "Second-order statistics of complex signals," *IEEE Trans. Signal Process.*, vol. 45, no. 2, pp. 411-420, Feb. 1997.
- [18] J. Yeo and J. H. Cho, "Asymptotic FRESH properizer for block processing of improper-complex second-order cyclostationary random processes," submitted to *IEEE Trans. Inf. Theory*, Apr. 2013, also available online at <http://arxiv.org/abs/1304.7375>.
- [19] E. K. P. Chong and S. H. Žak, *An Introduction to Optimization*, 4th ed. NJ: Wiley, 2013.

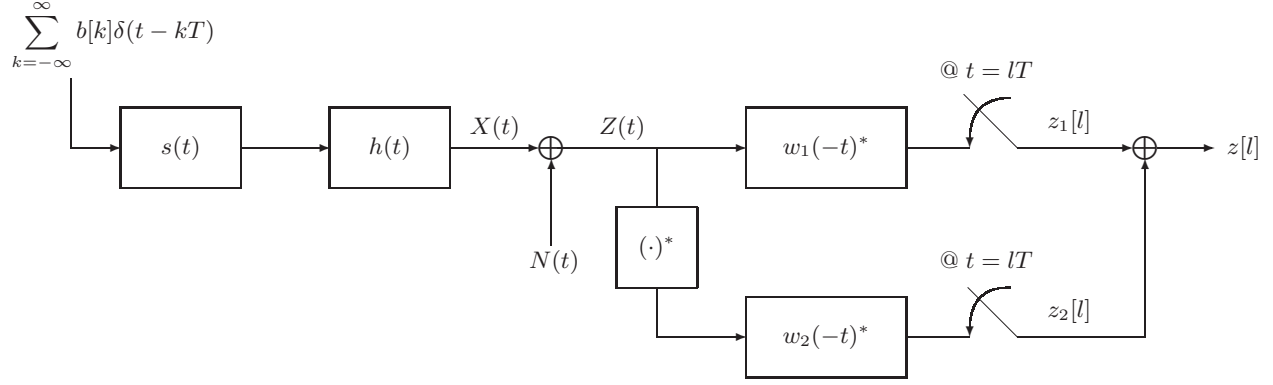


Fig. 1. System block diagram.

TABLE I
AN ALGORITHM TO CONSTRUCT CANDIDATE DENSITY FUNCTION AT $\nu \in (0, \nu_{\max}]$

-
- 1: Choose $f_0 \in \mathcal{F}^+$.
 - 2: Construct $a(f_0)$ and $a(-f_0)$ as follows.
 - 3: Set $a(-f_0) := 0$.
 - 4: **REPEAT**
 - 5: Update $a(f_0)$ as $a(f_0) := u(\nu, f_0)$ by using $u(\nu, f)$ defined in (26).
 - 6: Update $a(-f_0)$ as $a(-f_0) := u(\nu, -f_0)$.
 - 7: **UNTIL** $a(f_0)$ and $a(-f_0)$ converge.
 - 8: Repeat lines 1 – 7 for all $f_0 \in \mathcal{F}^+$.
-

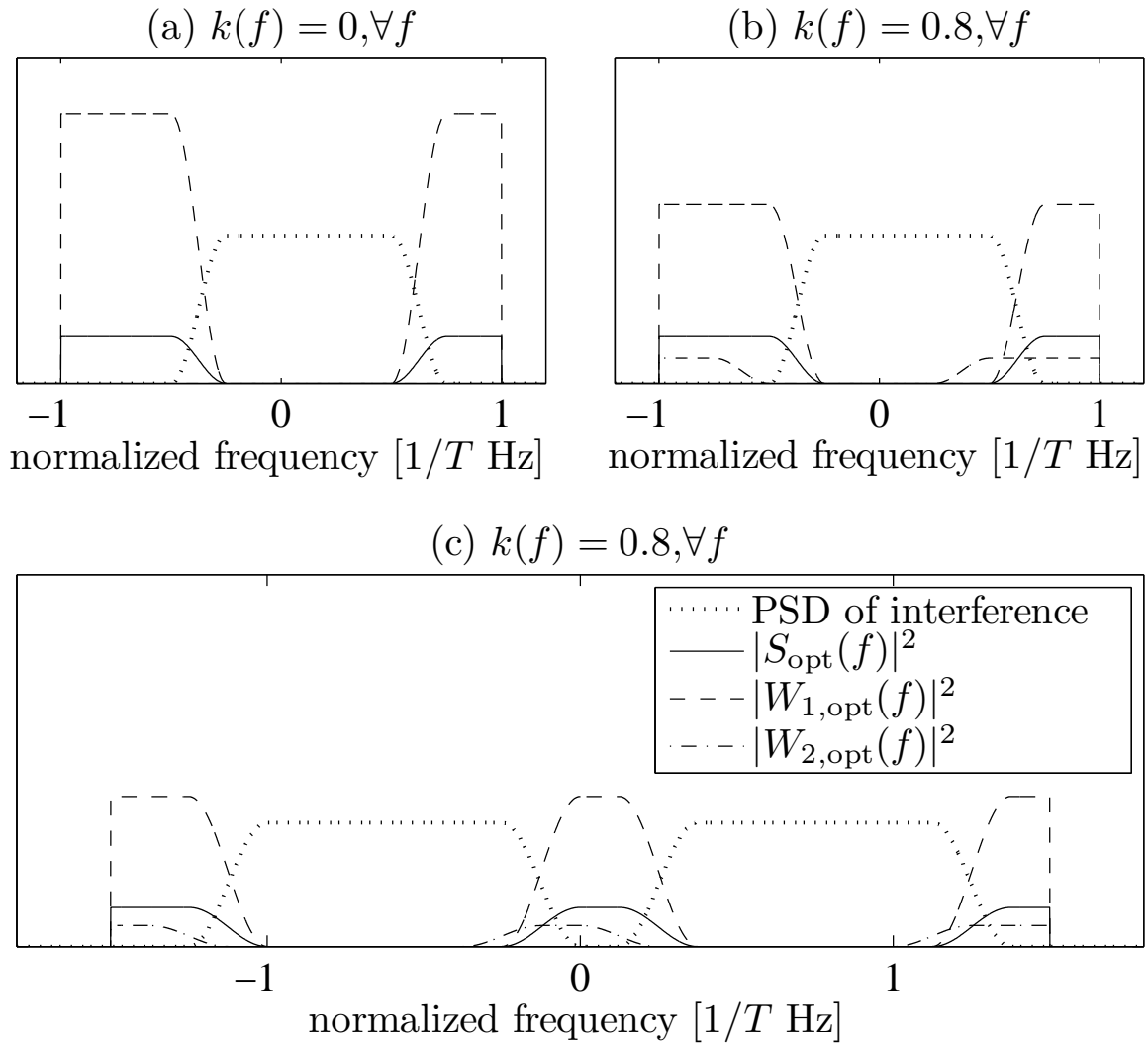


Fig. 2. Comparison of squared-magnitudes of the optimal transmit and receive waveforms for (a) $k(f) = 0, \forall f$ and single interferer, (b) $k(f) = 0.8, \forall f$ and single interferer, and (c) $k(f) = 0.8, \forall f$ and two uncorrelated interferers.

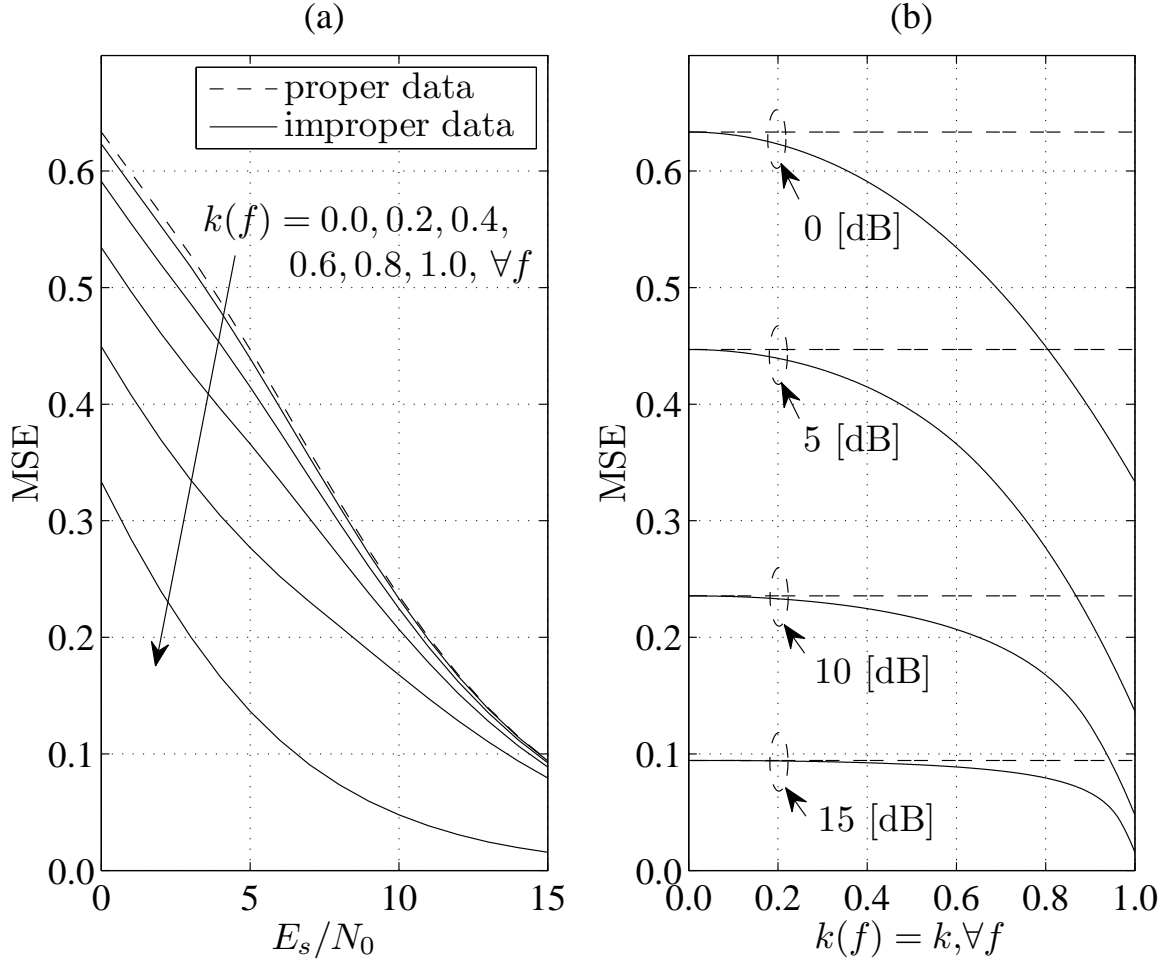


Fig. 3. Comparison of MSE (a) versus E_s/N_0 and (b) versus impropriety $k(f) = k, \forall f$.